

Sensitivity and Covariance in Stochastic Complementarity Problems with an Application to Natural Gas Markets

Sriram Sankaranarayanan¹

Felipe Feijoo²

Sauleh Siddiqui^{1,3,4}

¹ Department of Civil Engineering, Johns Hopkins University

² Joint Global Change Research Institute, Pacific Northwest National Laboratory

³ Department of Applied Mathematics and Statistics, Johns Hopkins University

⁴ German Institute for Economic Research (DIW Berlin)

Abstract

We provide an efficient method to approximate the covariance between decision variables and uncertain parameters in solutions to a general class of stochastic nonlinear complementarity problems. We also develop a sensitivity metric to quantify the uncertainty propagation in the problem by determining the change in the variance of the output variables due to a change in the variance of an input parameter. The covariance matrix of the solution variables quantifies the uncertainty in the output and pairs correlated variables and parameters. The sensitivity metric helps in identifying the parameters that cause maximum fluctuations in the output. The method developed in this paper optimizes the use of gradients and matrix multiplications which makes it particularly useful for large-scale problems. Having developed this method, we extend the deterministic version of the North American Natural Gas Model (NANGAM) to incorporate effects due to uncertainty in the parameters of the demand function, supply function, tariffs, and investment costs. We then use the sensitivity metrics to identify the parameters that impact the equilibrium the most.

1 Introduction

Complementarity problems arise naturally out of various real life problems. A rigorous survey of their application is available in [20]. [21, 35, 9, 28, 19, 37, 1, 11] use complementarity problems to model markets from a game theoretic perspective ([41, 4]), where the complementarity conditions typically arise between the marginal profit and the quantity produced by the producer. In the field of mechanics, they typically arise in the context of frictional contact problems ([32]), where there is a complementarity relation between the frictional force between a pair of surfaces and the distance of separation between them. [44] show the application of complementarity problems for pricing American options. With wide range of application, understanding the behavior of complementarity problems gains importance, especially in cases where there are uncertainties in the problem.

The behavior of a solution to a complementarity problem with random parameters was first addressed in [26], where such problems were referred to as stochastic complementarity problems (SCP). [39] formally defines two primary formulations of SCPs, namely the almost-sure formulation and the expectation based formulation. While the former formulation rarely has a solution, the latter is widely accepted as the SCP. In addition to this, [10] define an expected residual minimization (ERM) formulation for an SCP which they solve using Quasi-Monte Carlo methods. There have been a number of different methods employed to solve SCPs which include scenario-based methods, gradient-based methods and Monte-Carlo sampling methods. [23] and [15] use a scenario reduction based approach to solve the SCP, which systematically analyzes the probability of a discrete set of scenarios. Not restricting to discrete distributions for the random variables, [30] provide an iterative line-search based algorithm to converge to a solution of the SCP under assumptions of monotonicity.

While the scenario reduction methods typically assume discrete probability distributions for random parameters, it is more appropriate to sample certain real-life quantities from a continuous distribution. For example, the Central Limit Theorem and the Extreme Value Theorem guarantee that processes arising as sum of sufficiently many random variables and as maxima and minima of sufficiently many

variables follow Normal or one of the extreme value distributions respectively ([5, 24]). The ERM formulation and the solution by [10] solves the problem of finding the mean of the solution irrespective of the distribution of the parameters. While [33] compute confidence intervals for solution of the expected value formulation of the problem, we do not have efficient methods to find the second-order statistics for large-scale complementarity problems, *i.e.*, the covariance of the solution for problems over 10,000 output variables and over 1000 random input parameters.

Large-scale problems arise naturally out of detailed market models and there is considerable interest in studying, understanding and solving such models. For example, [6] and [43] discuss a case of routing and staffing in service systems where large-scale optimization is used. [22] discuss a case of an energy model with large number of variables and parameters. Naturally, developing methods to solve such large-scale problems gained interest. [12, 34, 8] discuss various tools ranging from mathematical techniques (Bender’s decomposition) to computational techniques (parallel processing) for solving large-scale optimization problems. [42] use an approximation approach to solve a loan portfolio selection problem.

The objective of this paper is to efficiently obtain second-order statistical information about solution vectors of large-scale stochastic complementarity problems. In addition, we also introduce a sensitivity metric which quantifies the change in uncertainty in the output due to a perturbation in the variance of uncertain parameters.

The diagonal elements of the covariance matrix, *i.e.*, the variance of the solution vector, quantify the uncertainty in the decision variables while the off-diagonal elements capture the linear dependency among pairs of decision variables. [29] and [7] provide examples in the area of clinical pathways and ecology respectively about the utility of understanding the variance of the solution in addition to the mean. They also show that a knowledge of variance aids better understanding and planning of the system. [3] emphasize the necessity to understand covariance as a whole rather than individual variances by quantifying “the loss incurred on ignoring correlations” in a stochastic programming model.

The sensitivity metric developed in this paper quantifies the sensitivity of the output uncertainty and thus helps us to directly compare input parameters by the amount of uncertainty they propagate to the solution.

In attaining the above objectives, apart from solving the stochastic complementarity problem in its expected value formulation, the most computationally expensive step is to solve a system of linear equations. We choose approximation methods over analytical methods, integration or Monte Carlo simulation because of the computational hurdle involved while implementing those methods for large-scale problems. The method we describe in this paper achieves the following:

- The most expensive step has to be performed just once, irrespective of the covariance of the input parameters. Once the linear system of equations is solved, for each given covariance scenario, we only perform two matrix multiplications.
- Approximating the covariance matrix and getting a sensitivity metric can be obtained by solving the said linear system just once.

The methods developed in this paper can also be used for a general differentiable optimization problems with linear equality constraints. We prove stronger results on error bounds for special cases of quadratic programming.

Having developed this method, we apply it to a large-scale stochastic natural gas model for North America, an extension of the deterministic model developed by [19] and determine the covariance of the solution variables. We then proceed to identify the parameters which have the greatest impact on the solution. A Python class for efficiently storing and operating on sparse arrays of dimension greater than two is created. This is useful for working with high-dimensional problems which have an inherent sparse structure in the gradients.

We divide the paper as follows. Section 2 formulates the problem and mentions the assumptions used in the paper. It then discusses the method used to solve the stochastic complementarity problem, develops the algorithm used to approximate the solution covariance and provides proofs for bounding the error. Section 3 develops framework to quantify the sensitivity of the solution to each of the random variables. Section 4 shows how the result can be applied for certain optimization problems with equality constraints. Having obtained the theoretical results, section 5 gives an example of an oligopoly where this method can be applied. The sub-section 5.2 compares the computational time of the method developed in this paper with a Monte-Carlo method showing the performance of the method for large-scale problems. Section

6 describes the Natural Gas Model to which the said method is applied. It also discusses the scenario we used on the model and the results hence obtained. Section 7 discusses the possible enhancements for the model and its limitations in the current form. Appendix A has the mathematical formulation of the natural gas model to which the method is applied and appendix B briefs about the Python class developed to handle the sparse `ndarray`.

2 Approximation of covariance

For the rest of the paper, all bold quantities are vectors. A subscript i for those quantities refer to the i -th component of the vector in Cartesian representation.

2.1 Stochastic Complementarity problem

We define a complementarity problem and a stochastic complementarity problem which are central to the results obtained in this paper. We use a general definition of complementarity problems and stochastic complementarity problems as stated below.

Definition 1. ([18]) Given $\mathbf{F} : \mathbb{R}^{n \times m} \mapsto \mathbb{R}^n$, and parameters $\theta \in \mathbb{R}^m$, the *parametrized complementarity problem* is to find $\mathbf{x} \in \mathbb{R}^n$ such that

$$\mathbb{K} \ni \mathbf{x} \perp \mathbf{F}(\mathbf{x}; \theta) \in \mathbb{K}^* \quad (2.1)$$

where \mathbb{K}^* , the dual cone of \mathbb{K} is defined as

$$\mathbb{K}^* = \{ \mathbf{x} \in \mathbb{R}^n : \mathbf{v}^T \mathbf{x} \geq 0 \quad \forall \quad \mathbf{v} \in \mathbb{K} \} \quad (2.2)$$

Definition 2. Given a cone $\mathbb{K} \in \mathbb{R}^n$ a random function $\mathbf{F} : \mathbb{K} \times \Omega \mapsto \mathbb{R}^n$, the stochastic complementarity problem (SCP) is to find $x \in \mathbb{R}^n$ such that

$$\mathbb{K} \ni \mathbf{x} \perp \mathbb{E} \mathbf{F}(\mathbf{x}; \omega) \in \mathbb{K}^* \quad (2.3)$$

We now make assumptions on the form of \mathbb{K} in (2.1). This form of \mathbb{K} helps in establishing an equivalence between a complementarity problem and a minimization problem which is key to derive the approximation method in this paper.

Assumption 1. \mathbb{K} in (2.1) is a Cartesian product of half spaces and full spaces, *i.e.*, for some $\mathcal{I} \subseteq \{1, 2, \dots, n\}$

$$\mathbb{K} = \{ \mathbf{x} \in \mathbb{R}^n : \mathbf{x}_i \geq 0 \quad \text{if } i \in \mathcal{I} \} \quad (2.4)$$

We now propose a lemma about the form of the dual cone of \mathbb{K} to understand the special form that it has. This will eventually help us in converting the NCP into an unconstrained minimization problem.

Lemma 1. The dual cone \mathbb{K}^* of the set assumed in assumption 1 is

$$\mathbb{K}^* = \mathbb{K}' = \left\{ \mathbf{x} \in \mathbb{R}^n : \begin{array}{ll} \mathbf{x}_i \geq 0 & \text{if } i \in \mathcal{I} \\ \mathbf{x}_i = 0 & \text{if } i \notin \mathcal{I} \end{array} \right\} \quad (2.5)$$

Proof. Proof To show this, we first prove that every element in \mathbb{K}' indeed is in \mathbb{K}^* . And then we prove for every element $\mathbf{x} \notin \mathbb{K}'$, there exists some $\mathbf{v} \in \mathbb{K}$ such that $\mathbf{v}^T \mathbf{x} < 0$. Consider an arbitrary \mathbf{x} in \mathbb{K}' .

$$\mathbf{v}^T \mathbf{x} = \sum_{i=1}^n \mathbf{v}_i \mathbf{x}_i = \sum_{i \in \mathcal{I}} \mathbf{v}_i \mathbf{x}_i + \sum_{i \notin \mathcal{I}} \mathbf{v}_i \mathbf{x}_i \geq 0 \quad (2.6)$$

where the final inequality follows from the fact that each term in the first summation is individually non-negative and each term in the second summation is 0. Thus we have $\mathbb{K}' \subseteq \mathbb{K}^*$.

Now to show the reverse containment, suppose there is $\mathbf{x} \in \mathbb{R}^n; \mathbf{x} \notin \mathbb{K}'$. This means, we either have

1. at least one index $j \in \mathcal{I}$ such that $\mathbf{x}_j < 0$ or
2. at least one index $j \notin \mathcal{I}$ such that $\mathbf{x}_j \neq 0$

Now,

$$\mathbf{v}^T \mathbf{x} = \sum_{i=1}^n \mathbf{v}_i \mathbf{x}_i = \mathbf{v}_j \mathbf{x}_j + \sum_{i \neq j} \mathbf{v}_i \mathbf{x}_i \quad (2.7)$$

In the first case, choose \mathbf{v} such that $[\mathbf{v}]_i = 0$ for $i \neq j$ and $\mathbf{v}_j = 1$. Clearly $\mathbf{v} \in \mathbb{K}$ and for this choice of \mathbf{v} , the above sum is negative, showing $\mathbf{x} \notin \mathbb{K}^*$. In the second case, choose \mathbf{v} such that $[\mathbf{v}]_i = 0$ for $i \neq j$ and $\mathbf{v}_j = -\text{sgn}(\mathbf{x}_j)$. Clearly $\mathbf{v} \in \mathbb{K}$ and for this choice of \mathbf{v} , the above sum is negative, showing $\mathbf{x} \notin \mathbb{K}^*$. Thus we show $(\mathbb{K}')^c \subseteq (\mathbb{K}^*)^c$, which implies the reverse containment and completes the proof. \square

2.2 The approximation algorithm

We now define C-functions, which are central to pose the complementarity problem into an unconstrained optimization problem. The equivalent formulation as an unconstrained optimization problem assists us in developing the algorithm.

Definition 3. ([18]) A function $\psi : \mathbb{R}^2 \mapsto \mathbb{R}$ is a *C-function* when

$$\left. \begin{aligned} \psi(x, y) &= 0 \\ &\Leftrightarrow \\ x \geq 0 \quad y \geq 0 \quad xy &= 0 \end{aligned} \right\} \quad (2.8)$$

We consider the following commonly used C-functions.

$$\psi_{FB}(x, y) = \sqrt{x^2 + y^2} - x - y \quad (2.9)$$

$$\psi_{min}(x, y) = \min(x, y) \quad (2.10)$$

$$\psi_{Man}(x, y) = \zeta(|x - y|) - \zeta(|x|) - \zeta(|y|) \quad (2.11)$$

where $\zeta(x)$ is some strictly monotonic real-valued function with $\zeta(0) = 0$

Under our assumptions on \mathbb{K} , the following two theorems show the equivalence of the complementarity problem and an unconstrained minimization problem. The first theorem shows the necessary condition for a solution of the complementarity problem.

Theorem 1. Suppose assumption 1 holds. Then every solution $\mathbf{x}^*(\theta)$ of the parameterized complementarity problem in (2.1), is a global minimum of the following function $f(\mathbf{x}; \theta)$,

$$\Phi_i(\mathbf{x}, \theta; \mathbf{F}) = \begin{cases} \mathbf{F}_i(\mathbf{x}, \theta) & \text{if } i \notin \mathcal{I} \\ \psi_i(\mathbf{x}_i, \mathbf{F}_i(\mathbf{x}, \theta)) & \text{if } i \in \mathcal{I} \end{cases} \quad (2.12)$$

$$f(\mathbf{x}; \theta) = \frac{1}{2} \|\Phi(\mathbf{x}; \mathbf{F})\|_2^2 \quad (2.13)$$

with an objective value 0, for some set of not necessarily identical C-functions ψ_i .

Proof. Proof Since \mathbf{x}^* solves the problem, following from the requirement that $\mathbf{F}(\mathbf{x}^*) \in \mathbb{K}^*$ and lemma 1, if $i \notin \mathcal{I}$, $\mathbf{F}_i(\mathbf{x}^*(\theta); \theta) = 0$.

For $i \in \mathcal{I}$, $\mathbf{x}^* \in \mathbb{K} \Rightarrow \mathbf{x}^*_i \geq 0$ and $\mathbf{F}(\mathbf{x}^*) \in \mathbb{K}^* \Rightarrow \mathbf{F}_i(\mathbf{x}^*) \geq 0$. Also from the requirement $\mathbf{x}^{*T} \mathbf{F}(\mathbf{x}^*) = 0$, one of the above two quantities should vanish for each $i \in \mathcal{I}$. But C-functions are precisely functions that vanish when both their arguments are non-negative and of them equal zero. So $\psi_i(\mathbf{x}^*, \mathbf{F}_i(\mathbf{x}^*)) = 0$. Thus each coordinate of Φ is individually zero, which makes $f(\mathbf{x}^*)$ vanish, which is the smallest value f can take. Thus \mathbf{x}^* is a global minimum of f . \square

Now we show that the sufficient condition for a point to be a solution of the complementarity problem is that it should be a global minimum to the same unconstrained optimization problem as in Theorem 1.

Theorem 2. Suppose assumption 1 holds. If a solution to the problem in (2.1) exists and $\mathbf{x}^*(\theta)$ is an unconstrained global minimizer of $f(\mathbf{x}; \theta)$ defined in (2.13), then $\mathbf{x}^*(\theta)$ solves the complementarity problem in (2.1).

Proof. Proof Since a solution exists for the NCP, we know by Theorem 1 that the minimum value f can take is 0. Suppose we have $\mathbf{x}^* \in \mathbb{R}^n$ such that $f(\mathbf{x}^*; \theta) = 0$. Since f is sum of squares, this can happen only if each of the individual terms are zero. This means for $i \notin \mathcal{I}$, $\mathbf{F}_i(\mathbf{x}^*) = 0$.

Now since $\psi_i(\mathbf{x}_i, \mathbf{F}_i(\mathbf{x}^*)) = 0$ for $i \in \mathcal{I}$, we know $\mathbf{F}_i(\mathbf{x}^*) \geq 0$. This combined with the previous point implies $\mathbf{F}(\mathbf{x}^*) \in \mathbb{K}^*$.

Also from the fact that $\psi_i(\mathbf{x}_i^*; \mathbf{F}_i(\mathbf{x}^*; \theta)) = 0$ for $i \in \mathcal{I}$, we know that $\mathbf{x}^* \in \mathbb{K}$. It also implies that $\mathbf{x}_i^* \mathbf{F}_i(\mathbf{x}^*) = 0$ for $i \in \mathcal{I}$. Thus

$$\mathbf{x}^{*T} \mathbf{F}(\mathbf{x}^*; \theta) = \sum_{i=1}^n \mathbf{x}_i^* \mathbf{F}_i \quad (2.14)$$

$$= \sum_{i \in \mathcal{I}} \mathbf{x}_i^* \mathbf{F}_i + \sum_{i \notin \mathcal{I}} \mathbf{x}_i^* \mathbf{F}_i \quad (2.15)$$

$$= 0 + 0 = 0 \quad (2.16)$$

This implies $\mathbf{x}^*(\theta) \perp \mathbf{F}(\mathbf{x}^*; \theta)$ and $\mathbf{x}^*(\theta)$ solves the complementarity problem. \square

We state the below corollary explicitly to show that the equivalence holds for a stochastic complementarity problem too.

Corollary 1. A solution \mathbf{x}^* of a stochastic complementarity problem with $\hat{\mathbf{F}} \equiv \mathbb{E}\mathbf{F}(\mathbf{x}; \theta(\omega))$ is a global minimum of the function f defined in (2.13).

Proof. Proof The proof is immediate by considering a new function $\tilde{F} : \mathbb{R}^n \mapsto \mathbb{R}^n$ such that $\tilde{F}(\mathbf{x}) = \mathbb{E}\mathbf{F}(\mathbf{x}; \theta(\omega))$. \square

Now given a function \mathbf{F} , and a set \mathbb{K} which satisfies assumption 1, and a solution of the NCP $\mathbf{x}^*(\hat{\theta})$ for some fixed $\theta = \hat{\theta}$, we define a vector valued function $\Phi : \mathbb{R}^{n \times m} \mapsto \mathbb{R}^n$ component-wise as follows.

$$\Phi_i(\mathbf{x}, \theta; \mathbf{F}) = \begin{cases} \mathbf{F}_i(\mathbf{x}, \theta) & \text{if } i \notin \mathcal{I} \\ \psi^2(\mathbf{x}_i, \mathbf{F}_i(\mathbf{x}, \theta)) & \text{if } i \in \mathcal{Z} \\ \psi(\mathbf{x}_i, \mathbf{F}_i(\mathbf{x}, \theta)) & \text{otherwise} \end{cases} \quad (2.17)$$

$$f(\mathbf{x}; \theta) = \frac{1}{2} \|\Phi(\mathbf{x}; \mathbf{F})\|_2^2 \quad (2.18)$$

$$\mathcal{Z} = \left\{ i \in \mathbb{I} : \mathbf{x}_i^*(\hat{\theta}) = \mathbf{F}_i(\mathbf{x}^*(\hat{\theta}); \hat{\theta}) = 0 \right\} \quad (2.19)$$

Note that if ψ is a C-function, ψ^2 is also a C-function since $\psi^2 = 0 \iff \psi = 0$. We observe from Theorem 1 and 2 that minimizing $f(\mathbf{x}; \theta)$ over \mathbf{x} is equivalent to solving the NCP in (2.1).

Now we assume conditions on the smoothness of \mathbf{F} so that the solution to a perturbed problem is sufficiently close to the original solution.

Assumption 2. $\mathbf{F}(\mathbf{x}; \theta)$ is twice continuously differentiable in \mathbf{x} and θ over an open set containing \mathbb{K} .

Given that the rest of the analysis is on the function $f(\mathbf{x}; \theta)$ defined in (2.13), we prove that a sufficiently smooth \mathbf{F} and a suitable ψ ensure a sufficiently smooth f .

Theorem 3. With assumption 2 holding, we state $f(\mathbf{x}; \hat{\theta})$ defined as in (2.18) is a twice continuously differentiable at $f(\mathbf{x}; \hat{\theta}) = 0$ for any C function ψ provided

1. ψ is twice differentiable at $\{(a, b) \in \mathbb{R}^2 : \psi(a, b) = 0\} \setminus \{(0, 0)\}$ with a finite derivative and finite second derivative.
2. ψ vanishes sufficiently fast near the origin. i.e.,

$$\lim_{(a, b) \rightarrow (0, 0)} \psi^2(a, b) \frac{\partial^2 \psi(a, b)}{\partial a \partial b} = 0 \quad (2.20)$$

Proof. Proof Given that f is a sum of squares, it is sufficient to prove each term individually is twice continuously differentiable to prove the theorem. Also since we are only interested where f vanishes, it is sufficient to prove the above property for each term where it vanishes.

Consider terms from $i \notin \mathcal{I}$. Since \mathbf{F}_i is twice continuously differentiable, \mathbf{F}_i^2 is twice continuously differentiable too.

Consider the case $i \in \mathcal{I}$; $\mathbf{x}_i^*(\hat{\theta}) = \mathbf{F}_i(\mathbf{x}_i^*(\hat{\theta}), \hat{\theta}) = 0$. These contribute a ψ^4 term to f . With the notation $\psi_i = \psi(\mathbf{x}_i, \mathbf{F}_i(\mathbf{x}; \theta))$, we clearly have,

$$\frac{\partial \psi_i^4}{\partial \mathbf{x}_j} = 4\psi_i^3 \left(\frac{\partial \psi_i}{\partial a} \delta_{ij} + \frac{\partial \psi_i}{\partial b} \frac{\partial \mathbf{F}_i}{\partial \mathbf{x}_j} \right) = 0 \quad (2.21)$$

$$\begin{aligned} \frac{\partial^2 \psi_i^4}{\partial \mathbf{x}_j \partial \mathbf{x}_k} &= 12\psi_i^2 \left(\frac{\partial \psi_i}{\partial a} \delta_{ij} + \frac{\partial \psi_i}{\partial b} \frac{\partial \mathbf{F}_i}{\partial \mathbf{x}_j} \right) \left(\frac{\partial \psi_i}{\partial a} \delta_{ik} + \frac{\partial \psi_i}{\partial b} \frac{\partial \mathbf{F}_i}{\partial \mathbf{x}_k} \right) \\ &\quad + 4\psi_i^3 \left(\frac{\partial^2 \psi_i}{\partial a^2} \delta_{ij} \delta_{ik} + \frac{\partial^2 \psi_i}{\partial a \partial b} \frac{\partial \mathbf{F}_i}{\partial \mathbf{x}_k} + \text{other terms} \right) \end{aligned} \quad (2.22)$$

$$= 0 \quad (2.23)$$

For the third case, we have

$$\frac{\partial \psi_i^2}{\partial \mathbf{x}_j} = 2\psi_i \left(\frac{\partial \psi_i}{\partial a} \delta_{ij} + \frac{\partial \psi_i}{\partial b} \frac{\partial \mathbf{F}_i}{\partial \mathbf{x}_j} \right) = 0 \quad (2.24)$$

$$\begin{aligned} \frac{\partial^2 \psi_i^2}{\partial \mathbf{x}_j \partial \mathbf{x}_k} &= \left(\frac{\partial \psi_i}{\partial a} \delta_{ij} + \frac{\partial \psi_i}{\partial b} \frac{\partial \mathbf{F}_i}{\partial \mathbf{x}_j} \right) \\ &\quad + 2\psi_i \left(\frac{\partial^2 \psi_i}{\partial a^2} \delta_{ij} \delta_{ik} + \frac{\partial^2 \psi_i}{\partial a \partial b} \frac{\partial \mathbf{F}_i}{\partial \mathbf{x}_k} + \text{other terms} \right) \end{aligned} \quad (2.25)$$

$$= \frac{\partial \psi_i}{\partial a} \delta_{ij} + \frac{\partial \psi_i}{\partial b} \frac{\partial \mathbf{F}_i}{\partial \mathbf{x}_j} \quad (2.26)$$

Continuity of f at the points of interest follow the continuity of the individual terms at the points. \square

The following corollaries show the existence of C-functions ψ which satisfy the hypothesis of Theorem 3.

Corollary 2. With assumption 2 holding, for the choice of C-function $\psi = \psi_{\min}$ defined in (2.10), the function f is twice continuously differentiable at its zeros.

Proof. Proof The set $O = \{(a, b) : \psi_{\min}(a, b) = 0\} \setminus \{(0, 0)\}$ is the positive coordinate axes except the origin. Rewriting this C-function as

$$\psi_{\min}(a, b) = \begin{cases} a & \text{if } a < b \\ b & \text{otherwise} \end{cases} \quad (2.27)$$

we see that the second derivative of ψ_{\min} vanishes at O . Also we observe that

$$\lim_{(a,b) \rightarrow (0,0)} \psi^2(a, b) \frac{\partial^2 \psi(a, b)}{\partial a \partial b} = 0 \quad (2.28)$$

since the second derivative is 0 everywhere except along the line $a = b$. \square

The following corollary gives a C-function which makes f twice continuously differentiable everywhere.

Corollary 3. With assumption 2 holding, for the choice of C-function $\psi = \psi_{FB}$ defined in (2.9), the function f is twice continuously differentiable.

Proof. Proof For $\psi = \psi_{FB}$, we have assumption 1 of Theorem 3 satisfied by [18]. For assumption 2,

$$\lim_{(a,b) \rightarrow (0,0)} \psi^2(a, b) \frac{\partial^2 \psi(a, b)}{\partial a \partial b} = \lim_{(a,b) \rightarrow (0,0)} \left(\sqrt{a^2 + b^2} - a - b \right)^2 \frac{ab}{(\sqrt{a^2 + b^2})^3} \quad (2.29)$$

$$= 0 \quad (2.30)$$

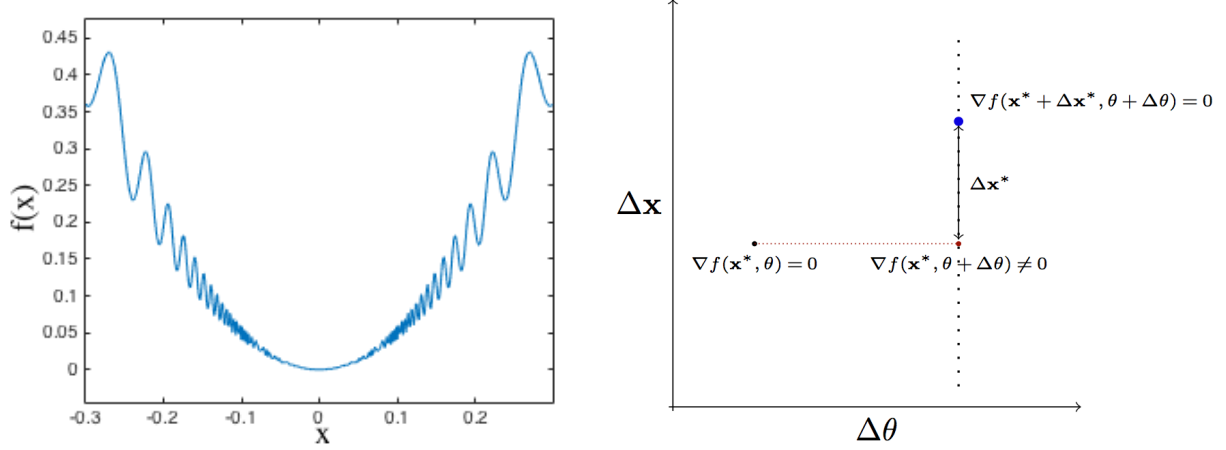


Figure 1: Left: An example of a function where the global minimum $x = 0$ is a non-isolated solution. Right: The intuition behind our approximation for finding where $\nabla f(\mathbf{x}, \theta) = 0$ under a small perturbation.

Thus f is twice continuously differentiable at its zeros. The twice continuous differentiability elsewhere follows directly from the fact that ψ_{FB} is twice continuously differentiable everywhere except at the origin. This ensures that all the terms in the derivative of the sum of squares exist and are finite. \square

We now define an isolated solution to a problem and assume that the problem of interest has this property. This is required to ensure that our approximation is well defined.

Definition 4. ([36]) A solution \mathbf{x}^* of a problem is said to be an *isolated solution*, if there is a neighborhood $\mathcal{B}(\mathbf{x}^*; \epsilon)$ of \mathbf{x}^* , where \mathbf{x}^* is the only solution of the problem.

A counter-example for an isolated minimum is shown on the left of Fig. 1. It is a plot of the function

$$f(x) = 5x^2 + x^2 \sin\left(\frac{1}{x^2}\right) \quad (2.31)$$

and the global minimum at $x = 0$ is not an isolated minimum as we can confirm that any open interval around $x = 0$ has other minimum contained in it. Unlike this case, in this paper, we assume that if we obtain a global-minimum of f , then it is an isolated minimum.

Assumption 3. For some fixed value of $\theta = \hat{\theta}$, there exists a known solution $\mathbf{x}^*(\hat{\theta})$ such that it is an isolated global minimum of $f(\mathbf{x}; \theta)$.

We propose algorithm 1 to approximate the covariance of the output given a covariance matrix for the input parameters. Following that Theorem 4 gives a mathematical proof that the algorithm indeed approximates the covariance matrix. Then Theorem 5 bounds the error in the approximation. The intuition behind the approximation is shown on the right of Fig. 1.

Theorem 4. Algorithm 1 generates a first-order approximation for the change in solution for a perturbation in parameters and computes the covariance of the solution for a complementarity problem with uncertain parameters with small variances.

Proof. Proof Consider the function $f(\mathbf{x}; \theta)$. From Theorem 1, \mathbf{x}^* minimizes this function for $\theta = \hat{\theta}$. From Theorem 3, we have $f(\mathbf{x}; \theta)$ is twice continuously differentiable at all its zeros. Thus we have,

$$\nabla_{\mathbf{x}} f(\mathbf{x}^*; \hat{\theta}) = 0 \quad (2.35)$$

Algorithm 1 Approximating Covariance

Require: Solve the complementarity problem in (2.1) for the mean value of $\theta = \hat{\theta}$, or solve the stochastic complementarity problem in (2.3) and calibrate the value of the parameters $\theta = \hat{\theta}$ for this solution. Call this solution as \mathbf{x}^* . Choose a tolerance level τ .

- 1: Evaluate $\mathbf{F}^* \leftarrow \mathbf{F}(\mathbf{x}^*; \hat{\theta})$, $G_{ij} \leftarrow \frac{\partial \mathbf{F}_i(\mathbf{x}^*; \hat{\theta})}{\partial \mathbf{x}_j}$, $L_{i,j} \leftarrow \frac{\partial \mathbf{F}_i(\mathbf{x}^*; \hat{\theta})}{\partial \theta_j}$.
- 2: Choose a C-function ψ such that the conditions in Theorem 3 are satisfied.
- 3: Define the function $\psi^a(a, b) = \frac{\partial \psi(a, b)}{\partial a}$, $\psi^b(a, b) = \frac{\partial \psi(a, b)}{\partial b}$.
- 4: Find the set of indices $\mathcal{Z} = \{z \in \mathcal{I} : |\mathbf{x}^*_z| = |\mathbf{F}^*_z| \leq \tau\}$.
- 5: Define

$$\mathcal{M}_{ij} \leftarrow \begin{cases} G_{ij} & \text{if } i \notin \mathcal{I} \\ 0 & \text{if } i \in \mathcal{Z} \\ \psi^a(\mathbf{x}^*_i, \mathbf{F}^*_i) \delta_{ij} + \psi^b(\mathbf{x}^*_i, \mathbf{F}^*_i) G_{ij} & \text{otherwise} \end{cases} \quad (2.32)$$

where $\delta_{ij} = 1$ if $i = j$ and 0 otherwise.

- 6: Define

$$\mathcal{N}_{ij} \leftarrow \begin{cases} L_{ij} & \text{if } i \notin \mathcal{I} \\ 0 & \text{if } i \in \mathcal{Z} \\ \psi^b(\mathbf{x}^*_i, \mathbf{F}^*_i) L_{ij} & \text{otherwise} \end{cases} \quad (2.33)$$

- 7: Solve the linear systems of equations for \mathcal{T} .

$$\mathcal{M}\mathcal{T} = \mathcal{N} \quad (2.34)$$

If \mathcal{M} is non singular, we have a unique solution. If not, a least square solution or a solution obtained by calculating the Moore Penrose Pseudo inverse ([27]) can be used.

- 8: Given \mathcal{C} , a covariance matrix of the input random parameters, $\theta(\omega)$, **return** $\mathcal{C}^* \leftarrow \mathcal{T}\mathcal{C}\mathcal{T}^T$.
-

Now suppose the parameters $\hat{\theta}$ are perturbed by $\Delta\theta$, then the above gradient can be written using the mean value theorem and then approximated up to the first order as follows.

$$\nabla_{\mathbf{x}} \mathbf{f}(\mathbf{x}^*(\hat{\theta}), \hat{\theta} + \Delta\theta) = \nabla_{\mathbf{x}} \mathbf{f}(\mathbf{x}^*(\hat{\theta}), \hat{\theta}) + \nabla_{\hat{\theta}} \nabla_{\mathbf{x}} \mathbf{f}(\mathbf{x}^*(\hat{\theta}), \tilde{\theta}) \Delta\theta \quad (2.36)$$

$$\nabla_{\mathbf{x}} \mathbf{f}(\mathbf{x}^*(\hat{\theta}), \hat{\theta} + \Delta\theta) - \nabla_{\mathbf{x}} \mathbf{f}(\mathbf{x}^*(\hat{\theta}), \hat{\theta}) \approx \mathcal{J} \Delta\theta \quad (2.37)$$

where,

$$\tilde{\theta} \in [\theta, \theta + \Delta\theta] \quad (2.38)$$

$$\mathcal{J}_{ij} = [\nabla_{\hat{\theta}} \nabla_{\mathbf{x}} \mathbf{f}(\mathbf{x}^*(\hat{\theta}), \hat{\theta})]_{ij} \quad (2.39)$$

$$= \frac{\partial [\nabla_{\mathbf{x}} \mathbf{f}(\mathbf{x}^*; \hat{\theta})]_i}{\partial \hat{\theta}_j} \quad (2.40)$$

Since $\mathcal{J} \Delta\theta$ is not guaranteed to be 0, we might have to alter \mathbf{x} to bring the gradient back to zero. *i.e.*, we need $\Delta\mathbf{x}$ such that $\nabla_{\mathbf{x}} \mathbf{f}(\mathbf{x}^*(\hat{\theta}) + \Delta\mathbf{x}, \hat{\theta} + \Delta\theta) = \mathbf{0}$. But by the mean value theorem,

$$\nabla_{\mathbf{x}} \mathbf{f}(\mathbf{x}^*(\hat{\theta}) + \Delta\mathbf{x}, \hat{\theta} + \Delta\theta) = \nabla_{\mathbf{x}} \mathbf{f}(\mathbf{x}^*(\hat{\theta}), \hat{\theta} + \Delta\theta) + \nabla_{\mathbf{x}}^2 \mathbf{f}(\tilde{\mathbf{x}}, \hat{\theta} + \Delta\theta) \Delta\mathbf{x} \quad (2.41)$$

$$\mathbf{0} \approx \mathcal{J} \Delta\theta + \nabla_{\mathbf{x}}^2 \mathbf{f}(\tilde{\mathbf{x}}, \hat{\theta}) \Delta\mathbf{x} \quad (2.42)$$

$$\approx \mathcal{J} \Delta\theta + \nabla_{\mathbf{x}}^2 \mathbf{f}(\mathbf{x}^*(\hat{\theta}), \hat{\theta}) \Delta\mathbf{x} \quad (2.43)$$

$$\mathcal{H} \Delta\mathbf{x} \approx -\mathcal{J} \Delta\theta \quad (2.44)$$

where,

$$\tilde{\mathbf{x}} \in [\mathbf{x}^*(\hat{\theta}), \mathbf{x}^*(\hat{\theta}) + \Delta\mathbf{x}] \quad (2.45)$$

$$[\mathcal{H}]_{ij} = [\nabla_{\mathbf{x}}^2 \mathbf{f}(\mathbf{x}^*(\hat{\theta}), \hat{\theta})]_{ij} \quad (2.46)$$

$$= \frac{\partial [\nabla_{\mathbf{x}} \mathbf{f}(\mathbf{x}^*; \hat{\theta})]_i}{\partial \mathbf{x}_j} \quad (2.47)$$

Now from [36], the gradient of the least squares function f can be written as

$$\nabla_{\mathbf{x}} f(\mathbf{x}^*, \hat{\theta}) = \mathcal{M}^T \Phi(\mathbf{x}^*, \hat{\theta}) \quad (2.48)$$

$$[\mathcal{M}]_{ij} = \frac{\partial \Phi_i(\mathbf{x}^*, \hat{\theta})}{\partial \mathbf{x}_j} \quad (2.49)$$

$$= \begin{cases} \frac{\partial \mathbf{F}_i(\mathbf{x}^*, \hat{\theta})}{\partial \mathbf{x}_j} & \text{if } i \notin \mathcal{I} \\ \frac{\partial \psi^2(\mathbf{x}_i, \mathbf{F}_i(\mathbf{x}^*, \hat{\theta}))}{\partial \mathbf{x}_j} & \text{if } i \in \mathcal{Z} \\ \frac{\partial \psi(\mathbf{x}_i, \mathbf{F}_i(\mathbf{x}^*, \hat{\theta}))}{\partial \mathbf{x}_j} & \text{otherwise} \end{cases} \quad (2.50)$$

$$= \begin{cases} \frac{\partial \mathbf{F}_i(\mathbf{x}^*, \hat{\theta})}{\partial \mathbf{x}_j} & \text{if } i \notin \mathcal{I} \\ 0 & \text{if } i \in \mathcal{Z} \\ \frac{\partial \psi_i}{\partial \mathbf{x}_j} & \text{otherwise} \end{cases} \quad (2.51)$$

which is the form of \mathcal{M} defined in algorithm 1. Also

$$\mathcal{H} = \nabla_{\mathbf{x}}^2 f(\mathbf{x}^*; \hat{\theta}) = \mathcal{M}^T \mathcal{M} + \sum_{i=1}^n \Phi_i(\mathbf{x}^*; \hat{\theta}) \nabla_{\mathbf{x}}^2 \Phi_i(\mathbf{x}^*; \hat{\theta}) \quad (2.52)$$

$$= \mathcal{M}^T \mathcal{M} \quad (2.53)$$

where the second term vanishes since we have from Theorem 1 that each term of Φ individually vanishes at the solution. Now

$$\mathcal{J} = \nabla_{\mathbf{x}\theta} f(\mathbf{x}^*; \hat{\theta}) \quad (2.54)$$

$$\mathcal{J}_{ij} = \frac{\partial [\nabla_{\mathbf{x}} f(\mathbf{x}^*; \hat{\theta})]_i}{\partial \theta_j} \quad (2.55)$$

$$= \frac{\partial}{\partial \theta_j} \left(\sum_{k=1}^n [\nabla_{\mathbf{x}} \Phi(\mathbf{x}^*; \hat{\theta})]_{ki} \Phi_k(\mathbf{x}^*; \hat{\theta}) \right) \quad (2.56)$$

$$= \sum_{k=1}^n \left(\frac{\partial [\nabla_{\mathbf{x}} \Phi(\mathbf{x}^*; \hat{\theta})]_{ki}}{\partial \theta_j} \Phi_k(\mathbf{x}^*; \hat{\theta}) + [\nabla_{\mathbf{x}} \Phi(\mathbf{x}^*; \hat{\theta})]_{ki} \frac{\partial \Phi_k(\mathbf{x}^*; \hat{\theta})}{\partial \theta_j} \right) \quad (2.57)$$

$$= \sum_{k=1}^n \mathcal{M}_{ki} \mathcal{N}_{kj} = \mathcal{M}^T \mathcal{N} \quad (2.58)$$

where the first term vanished because Φ_i are individually zeros, and we define

$$\mathcal{N}_{ij} = \frac{\partial \Phi_k(\mathbf{x}^*; \hat{\theta})}{\partial \theta_j} \quad (2.59)$$

$$= \begin{cases} \frac{\partial \mathbf{F}_i}{\partial \mathbf{x}_j} & \text{if } i \notin \mathcal{I} \\ 2\psi(\mathbf{x}^*_i; \mathbf{F}_i^*) \psi^b(\mathbf{x}^*_i; \mathbf{F}_i^*) \frac{\partial \mathbf{F}_i}{\partial \mathbf{x}_j} & \text{if } i \in \mathcal{Z} \\ \psi^b(\mathbf{x}^*_i; \mathbf{F}_i^*) \frac{\partial \mathbf{F}_i(\mathbf{x}^*; \hat{\theta})}{\partial \mathbf{x}_j} & \text{otherwise} \end{cases} \quad (2.60)$$

$$= \begin{cases} \frac{\partial \mathbf{F}_i}{\partial \mathbf{x}_j} & \text{if } i \notin \mathcal{I} \\ 0 & \text{if } i \in \mathcal{Z} \\ \psi^b(\mathbf{x}^*_i; \mathbf{F}_i^*) \frac{\partial \mathbf{F}_i(\mathbf{x}^*; \hat{\theta})}{\partial \mathbf{x}_j} & \text{otherwise} \end{cases} \quad (2.61)$$

which is the form of \mathcal{N} defined in algorithm 1. By assumption 3, we have a unique minimum in the neighborhood of \mathbf{x}^* where the gradient vanishes. So we have from (2.44), (2.53) and (2.58)

$$\mathcal{H} \Delta \mathbf{x} = -\mathcal{J} \Delta \theta \quad (2.62)$$

$$\mathcal{M}^T \mathcal{M} \Delta \mathbf{x} = -\mathcal{M}^T \mathcal{N} \Delta \theta \quad (2.63)$$

$\Delta \mathbf{x}$ solves the above equation, if it solves

$$\mathcal{M}\Delta \mathbf{x} = -\mathcal{N}\Delta \theta \quad (2.64)$$

By defining \mathcal{T} as the solution to the linear system of equations

$$\mathcal{M}\mathcal{T} = \mathcal{N} \quad (2.65)$$

$$\Delta \mathbf{x} = -\mathcal{T}\Delta \theta \quad (2.66)$$

and we have the above first-order approximation. From [38], we know that if some vector \mathbf{x} has covariance \mathbf{C} , then for a matrix A , the vector $A\mathbf{x}$ will have covariance ACA^T . So we have.

$$\text{Cov}(\Delta \mathbf{x}) \approx \mathcal{T} \text{Cov}(\Delta \theta) \mathcal{T}^T \quad (2.67)$$

This forms the basis for the computation of the covariance of $\Delta \mathbf{x}$ in algorithm 1. \square

For computational purposes the matrix \mathcal{T} in the above equation has to be calculated only once, irrespective of the number of scenarios for which we would like to run for the covariance of θ . Thus if $\mathbf{x} \in \mathbb{R}^n$, $\theta \in \mathbb{R}^m$ and we want to test the output covariance for k different input covariance cases, the complexity is equal to that of solving a system of n linear equations m times as in (2.65), and hence is $O(mn^2)$. *i.e.*, the complexity is quadratic in the number of output variables, linear in the number of input parameters and constant in the number of covariance scenarios we would like to run.

In theorem 5 below, we prove that the error in the approximation of theorem 4 can be bounded using the condition number of the Hessian. We need the following assumption that the condition number of the Hessian of \mathbf{f} is bounded and the Hessian is Lipschitz continuous.

Assumption 4. At the known solution of the complementarity problem of interest ($\theta = \hat{\theta}$),

1. The condition number of the Hessian of \mathbf{f} defined is finite and equal to κ_H
2. The Hessian of \mathbf{f} is Lipschitz continuous with a Lipschitz constant $\mathcal{L}(\mathbf{x}^*; \theta)$.

Theorem 5. With assumption 4 holding, the error in the linear approximation 2.44 for a perturbation of ϵ is $o(\epsilon)$.

Proof. Proof Since $\nabla^2 f$ is Lipschitz continuous on both \mathbf{x} and θ , we can write for $\tilde{\mathbf{x}}$ near \mathbf{x}^* ,

$$\left\| \nabla_{\mathbf{x}}^2 f(\mathbf{x}^*(\hat{\theta}), \hat{\theta}) - \nabla_{\mathbf{x}}^2 f(\tilde{\mathbf{x}}, \hat{\theta}) \right\| \leq \mathcal{L}(\mathbf{x}^*; \theta) \left\| \mathbf{x}^*(\hat{\theta}) - \tilde{\mathbf{x}} \right\| \quad (2.68)$$

$$\leq \mathcal{L}(\mathbf{x}^*; \theta) \|\Delta \mathbf{x}\| \quad (2.69)$$

$$\tilde{\mathcal{H}} = \nabla_{\mathbf{x}}^2 f(\tilde{\mathbf{x}}, \hat{\theta}) \quad (2.70)$$

$$= \mathcal{H} + \varepsilon_H \quad (2.71)$$

where $\|\varepsilon_H\| \leq \mathcal{L}(\mathbf{x}^*; \theta) \|\Delta \mathbf{x}\|$. Applying the Lipschitz continuity on θ ,

$$\left\| \nabla_{\theta} \nabla_{\mathbf{x}} f(\mathbf{x}^*(\hat{\theta}), \hat{\theta}) - \nabla_{\theta} \nabla_{\mathbf{x}} f(\mathbf{x}^*(\hat{\theta}), \tilde{\theta}) \right\| \leq \mathcal{L}(\mathbf{x}^*; \theta) \left\| \tilde{\theta} - \hat{\theta} \right\| \quad (2.72)$$

$$\leq \mathcal{L}(\mathbf{x}^*; \theta) \|\Delta \theta\| \quad (2.73)$$

$$\tilde{\mathcal{J}} = \nabla_{\theta} \nabla_{\mathbf{x}} f(\mathbf{x}^*(\hat{\theta}), \tilde{\theta}) \quad (2.74)$$

$$= \mathcal{J} + \varepsilon_J \quad (2.75)$$

where $\|\varepsilon_J\| \leq \mathcal{L}(\mathbf{x}^*; \theta) \|\Delta \theta\|$. Thus the equation

$$\tilde{\mathcal{H}}\Delta \mathbf{x} = \tilde{\mathcal{J}}\Delta \theta \quad (2.76)$$

is exact, even if we cannot compute $\tilde{\mathcal{H}}$ and $\tilde{\mathcal{J}}$ exactly. Now the error in inverting $\tilde{\mathcal{H}}$ is bounded by the condition number ([27]).

$$\frac{\left\| \mathcal{H}^{-1} - \tilde{\mathcal{H}}^{-1} \right\|}{\left\| \tilde{\mathcal{H}}^{-1} \right\|} \leq \frac{\kappa_H \frac{\|\varepsilon_H\|}{\|\mathcal{H}\|}}{1 - \kappa_H \frac{\|\varepsilon_H\|}{\|\mathcal{H}\|}} \quad (2.77)$$

Assuming $\kappa_H \|\varepsilon_H\| \ll \|\mathcal{H}\|$, the above equation becomes

$$\frac{\|\mathcal{H}^{-1} - \tilde{\mathcal{H}}^{-1}\|}{\|\tilde{\mathcal{H}}^{-1}\|} \leq \kappa_H \frac{\|\varepsilon_H\|}{\|\tilde{\mathcal{H}}\|} \quad (2.78)$$

$$\Rightarrow \|\tilde{\mathcal{H}}^{-1} - \mathcal{H}^{-1}\| \leq \kappa_H \frac{\|\tilde{\mathcal{H}}^{-1}\|}{\|\tilde{\mathcal{H}}\|} \|\varepsilon_H\| \quad (2.79)$$

$$\Rightarrow \|\tilde{\mathcal{H}}^{-1} \tilde{\mathcal{J}} - \mathcal{H}^{-1} \tilde{\mathcal{J}} - \mathcal{H}^{-1} \varepsilon_J\| \leq \kappa_H \frac{\|\tilde{\mathcal{H}}^{-1}\|}{\|\tilde{\mathcal{H}}\|} \varepsilon_H \|\mathcal{J}\| + \kappa_H \frac{\|\tilde{\mathcal{H}}^{-1}\|}{\|\tilde{\mathcal{H}}\|} \|\varepsilon_H\| \|\varepsilon_J\| \quad (2.80)$$

$$\Rightarrow \|\tilde{\mathcal{H}}^{-1} \tilde{\mathcal{J}} - \mathcal{H}^{-1} \mathcal{J}\| \leq k_1 \|\Delta \mathbf{x}\| + k_2 \|\Delta \theta\| \quad (2.81)$$

with

$$k_1 = \kappa_H \mathcal{L}(\mathbf{x}^*; \theta) \frac{\|\tilde{\mathcal{H}}^{-1}\|}{\|\tilde{\mathcal{H}}\|} \|\mathcal{J}\| \quad (2.82)$$

$$k_2 = \mathcal{L}(\mathbf{x}^*; \theta) \|\mathcal{H}^{-1}\| \quad (2.83)$$

Thus we have from (2.81), that the error in the approximation done in algorithm 1 is bounded. \square

3 Stochastic Sensitivity Analyses

In this section, we quantify the sensitivity of the solution to each of the input parameters. To achieve this, we define total linear sensitivity. We then show how these quantities can be calculated using the matrix \mathcal{T} derived earlier. We then proceed to prove that these quantities also bound the maximum increase in uncertainties of the output. We now define a quantity that bounds a maximum change in the value of a function due to perturbations.

Definition 5. Given a function $\mathbf{f} : \mathbb{R}^m \mapsto \mathbb{R}^n$, the *total linear sensitivity*, $\beta_d \in \mathbb{R}_+$ of a dimension $d \leq m$; $d \in \mathbb{N}$ at a point $\mathbf{x} \in \mathbb{R}^m$ is defined for $\delta > 0$, sufficiently small,

$$\beta_d = \inf \{ \alpha : \|\mathbf{f}(\mathbf{x})\|_2 - \delta \alpha + o(\delta^2) \leq \|\mathbf{f}(\mathbf{x} + \delta e_d)\|_2 \leq \|\mathbf{f}(\mathbf{x})\|_2 + \delta \alpha + o(\delta^2) \} \quad (3.1)$$

where e_d is the d -th standard basis vector.

This is a bound to the distance by which the functional value can move for a small perturbation in the input. A solution to the parametrized complementarity problem in (2.1) can be viewed as a function from the space of parameter tuples to the solution of the problem. The above definition talks about bounding the change in the solution for a small change in the parameters. The next theorem shows how the total linear sensitivity can be calculated from the linear approximation matrix \mathcal{T} derived earlier.

Theorem 6. Suppose we know, $G \in \mathbb{R}^{n \times m}$ such that $G_{ij} = \frac{\partial f_i(\mathbf{x})}{\partial x_j}$, then $\beta_d = \sqrt{(\sum_{i=1}^n G_{id}^2)}$

Proof. Proof By definition, for some admissible d ,

$$\mathbf{f}(\mathbf{x} + \delta e_d) = \mathbf{f}(\mathbf{x}) + \delta G e_d + o(\delta^2) \quad (3.2)$$

$$\Rightarrow [\mathbf{f}(\mathbf{x} + \delta e_d)]_i = [\mathbf{f}(\mathbf{x})]_i + \delta G_{id} + o(\delta^2) \quad (3.3)$$

$$\|\mathbf{f}(\mathbf{x} + \delta e_d)\|_2 \leq \|\mathbf{f}(\mathbf{x})\|_2 + \|\delta G_{\cdot d}\|_2 + \|o(\delta^2)\|_2 \quad (3.4)$$

$$= \|\mathbf{f}(\mathbf{x})\|_2 + \delta \sqrt{\left(\sum_{i=1}^n G_{id}^2 \right)} + o(\delta^2) \quad (3.5)$$

where $G_{\cdot d}$ is the d -th column of G . Also we have from (3.3) for sufficiently small δ ,

$$\|\mathbf{f}(\mathbf{x} + \delta e_d)\|_2 \leq \|\mathbf{f}(\mathbf{x})\|_2 - \|\delta G_{\cdot d}\|_2 + o(\delta^2) \quad (3.6)$$

$$= \|\mathbf{f}(\mathbf{x})\|_2 - \delta \sqrt{\left(\sum_{i=1}^n G_{id}^2\right)} + o(\delta^2) \quad (3.7)$$

□

The above theorem proves that the \mathcal{T} matrix obtained in (2.65) is sufficient to approximate the total linear sensitivity. The following result suggests how the total linear sensitivity can approximate the total variance in the output variables.

Theorem 7. Given a function $\mathbf{f} : \mathbb{R}^m \mapsto \mathbb{R}^n$ and β_d , the increase in the total uncertainty in the output, *i.e.*, the sum of variances of the output variables, for a small increase of the variance of an input parameter, σ_d^2 of \mathbf{x}_d is approximated by $\beta_d^2 \sigma_d^2$.

Proof. Let E_d be the matrix of size $m \times m$ with zeros everywhere except the d -th diagonal element, where it is 1. Given $C = \text{Cov}(\mathbf{x}(\omega))$, for a small perturbation σ^2 in the variance of \mathbf{x}_d , the covariance of $\mathbf{f}(\mathbf{x})$ changes as follows.

$$C^* \approx \nabla_{\mathbf{x}} \mathbf{f} C \nabla_{\mathbf{x}} \mathbf{f}^T \quad (3.8)$$

$$C^* + \Delta C^* \approx \nabla_{\mathbf{x}} \mathbf{f} (C + \sigma^2 E_d) \nabla_{\mathbf{x}} \mathbf{f}^T \quad (3.9)$$

$$= C^* + \sigma^2 \nabla_{\mathbf{x}} \mathbf{f} E_d \nabla_{\mathbf{x}} \mathbf{f}^T \quad (3.10)$$

$$[\Delta C^*]_{ij} \approx \sigma^2 [\nabla_{\mathbf{x}} \mathbf{f}]_{id} [\nabla_{\mathbf{x}} \mathbf{f}]_{jd} \quad (3.11)$$

$$\sum_{i=1}^n [\Delta C^*]_{ii} \approx \sigma^2 \beta_d^2 \quad (3.12)$$

which is the total increase in variance. The off-diagonal terms do not affect the total uncertainty in the system because, the symmetric matrix C can be diagonalized as $Q D Q^T$, where Q is a rotation matrix, and the trace is invariant under orthogonal transformations. □

With the above result, we can determine the contribution of each input parameter to the total uncertainty in the output. Once the parameter which contributes the most to the output uncertainty is identified, efforts can be made to get a more accurate estimate of the parameter.

4 Application to unconstrained minimization

We now show how the method explained in algorithm 1 can be applied to approximate the covariance of the solution of an unconstrained optimization problem or an optimization problem with only equality constraints.

To start with, we assume conditions on the differentiability and convexity of the objective function.

Assumption 5. The objective function $\mathbf{f}(\mathbf{x}; \theta)$ is strictly convex in \mathbf{x} and is twice continuously differentiable in \mathbf{x} and θ .

In the theorem below, we approximate the covariance of the decision variables of a convex optimization with uncertainties in the linear term and with only linear equality constraints.

Theorem 8. With assumption 5 holding, the covariance of the primal and dual variables at the optimum of the problem,

$$\underset{\mathbf{x}}{\text{Minimize}} \quad \mathbf{f}(\mathbf{x}; \theta) = g(\mathbf{x}) + c(\theta)^T \mathbf{x} \quad (4.1)$$

$$\text{subject to } A\mathbf{x} = b(\theta) \quad (\mathbf{y}) \quad (4.2)$$

where $\theta = \theta(\omega)$ are random parameters with covariance C , is first-order approximated by $\mathcal{T}C\mathcal{T}^T$ where

$$\mathcal{T} = \begin{pmatrix} \nabla_{\mathbf{x}}^2 g(\mathbf{x}) & A^T \\ A & 0 \end{pmatrix}^{-1} \begin{pmatrix} -\nabla_{\theta} c(\theta) \\ \nabla_{\theta} b(\theta) \end{pmatrix} \quad (4.3)$$

Proof. Proof For the given optimization problem, because of assumption 5 and linear independence constraint qualification (LICQ), the KKT conditions are necessary and sufficient for optimality. The KKT condition satisfied at a solution $(\mathbf{x}^*, \mathbf{y}^*)$ for the problem are given by

$$\nabla_{\mathbf{x}} g(\mathbf{x}^*) + c(\theta) + A^T \mathbf{y}^* = 0 \quad (4.4)$$

$$A\mathbf{x}^* = b(\theta) \quad (4.5)$$

for some vector \mathbf{y} so that the equation is well defined. Suppose from there, θ is perturbed by $\Delta\theta$, we have

$$\nabla_{\mathbf{x}} g(\mathbf{x}^*) + c(\theta + \Delta\theta) + A^T \mathbf{y}^* \approx \nabla_{\theta} c(\theta) \Delta\theta \quad (4.6)$$

$$A\mathbf{x}^* - b(\theta + \Delta\theta) \approx -\nabla_{\theta} b(\theta) \Delta\theta \quad (4.7)$$

Now we need to find $\Delta\mathbf{x}$ and $\Delta\mathbf{y}$ such that

$$\nabla_{\mathbf{x}} g(\mathbf{x}^* + \Delta\mathbf{x}) + c(\theta + \Delta\theta) + A^T (\mathbf{y}^* + \Delta\mathbf{y}) \approx 0 \quad (4.8)$$

$$A(\mathbf{x}^* + \Delta\mathbf{x}) - b(\theta + \Delta\theta) \approx 0 \quad (4.9)$$

$$\nabla_{\mathbf{x}}^2 g(\mathbf{x}^*) \Delta\mathbf{x} + A^T \Delta\mathbf{y} \approx \nabla_{\theta} c(\theta) \Delta\theta \quad (4.10)$$

$$A \Delta\mathbf{x} \approx -\nabla_{\theta} b(\theta) \Delta\theta \quad (4.11)$$

The above conditions can be compactly represented as

$$\begin{pmatrix} \nabla_{\mathbf{x}}^2 g(\mathbf{x}) & A^T \\ A & 0 \end{pmatrix} \begin{pmatrix} \Delta\mathbf{x} \\ \Delta\mathbf{y} \end{pmatrix} = \begin{pmatrix} \nabla_{\theta} c(\theta) \\ -\nabla_{\theta} b(\theta) \end{pmatrix} \Delta\theta \quad (4.12)$$

If A has full rank, then the above matrix is non-singular. So the change in the decision variables \mathbf{x} and the duals \mathbf{y} can be written as a linear transformation of the perturbation in the random parameters. And we now have

$$\text{Cov} \begin{pmatrix} \Delta\mathbf{x} \\ \Delta\mathbf{y} \end{pmatrix} = \mathcal{T} \text{Cov}(\theta) \mathcal{T}^T \quad (4.13)$$

$$\mathcal{T} = \begin{pmatrix} \nabla_{\mathbf{x}}^2 g(\mathbf{x}) & A^T \\ A & 0 \end{pmatrix}^{-1} \begin{pmatrix} -\nabla_{\theta} c(\theta) \\ \nabla_{\theta} b(\theta) \end{pmatrix} \quad (4.14)$$

□

In the theorem below, we show that the method suggested is accurate (*i.e.*, has zero error) for an unconstrained quadratic optimization problem with uncertainty in the linear term.

Theorem 9. For an optimization problem with uncertainty of objectives of the form,

$$f(\mathbf{x}; \theta) = \frac{1}{2} \mathbf{x}^T G \mathbf{x} + \theta(\omega)^T \mathbf{x} \quad (4.15)$$

the approximation method has zero error. In other words, the obtained covariance matrix is exact.

Proof. Proof For the problem to be well-defined, let $G \in \mathbb{R}^{n \times n}$ and $\theta \in \mathbb{R}^n$. This makes $\nabla_{\mathbf{x}\theta}^2 f(\mathbf{x}; \theta) \in \mathbb{R}^{n \times n}$.

$$\nabla_{\mathbf{x}} f(\mathbf{x}; \theta) = G\mathbf{x} + \theta(\omega) \quad (4.16)$$

$$\nabla_{\mathbf{x}}^2 f(\mathbf{x}; \theta) = G \quad (4.17)$$

$$[\nabla_{\mathbf{x}\theta}^2 f(\mathbf{x}; \theta)]_{ij} = I \quad (4.18)$$

Due to absence of terms dependent on \mathbf{x} in the last two equations, we have an exact equation,

$$G\Delta\mathbf{x} = \Delta\theta \quad (4.19)$$

Due to the exactness of the above equation, we have

$$\mathcal{T} = G^{-1} \quad (4.20)$$

$$\text{Cov}(\Delta\mathbf{x}) = \mathcal{T} \text{Cov}(\Delta\theta) \mathcal{T}^T \quad (4.21)$$

with no error. \square

5 Application to a general oligopoly market

We now present an example of a complementarity problem in a natural gas oligopoly and show how the methods developed in this paper can be applied.

5.1 Problem Formulation and results

Consider k producers competitively producing natural gas in a Nash-Cournot game. Let us assume the unit costs of production are $\gamma_i(\omega)$, $i \in \{1, \dots, k\}$. We assume these are random variables. Also, let us assume that the consumer behavior is modeled by a linear demand curve $P(\tilde{Q})$ as follows.

$$P = a(\omega) + b(\omega)\tilde{Q} \quad (5.1)$$

where P is the price the consumer is willing to pay, \tilde{Q} is the total quantity of the natural gas produced and $a(\omega) > 0$, $b(\omega) < 0 \forall \omega \in \Omega$. Suppose the producers are competitively producing natural gas in a Nash-Cournot game, then the problem that each producer solves is as follows.

$$\text{Producer } i : \quad \text{Maximize} \left(a + b \left(\sum_{j=1}^k Q_j \right) \right) Q_i - \gamma_i Q_i \quad \text{s.t. } Q_i \geq 0 \quad (5.2)$$

where Q_i are the quantities produced by each producer. The KKT conditions of this optimization problem can be written as the following complementarity problem to obtain a Nash equilibrium ([13, 21]).

In this formulation, a, b, γ_i correspond to θ and Q_i correspond to \mathbf{x} in (2.1) with $\mathcal{I} = \{1, 2, \dots, k\}$.

$$\mathbf{F}_i(\mathbf{Q}) = \gamma_i - a - b \left(\sum_{j=1}^k Q_j \right) - b Q_i \quad (5.3)$$

In the current numerical example, let us consider a duopoly where $k = 2$. Let

$$\mathbb{E} \begin{pmatrix} \gamma_1 & \gamma_2 & a & b \end{pmatrix}^T = \begin{pmatrix} 2 & 1 & 15 & -1 \end{pmatrix}^T \quad (5.4)$$

Solving the complementarity problem deterministically with the above parameter values using the PATH algorithm ([14]), we get Q_1 and Q_2 to be 4 and 5 respectively. We use the C-function $\psi_{\min}(x, y) = \min(x, y)$ for this example to get

$$\mathcal{M} = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} \quad \mathcal{N} = \begin{pmatrix} 1 & 0 & -1 & -13 \\ 0 & 1 & -1 & -14 \end{pmatrix} \quad (5.5)$$

Now we have from (2.65)

$$\mathcal{T} = \mathcal{M}^{-1}\mathcal{N} = \frac{1}{3} \begin{pmatrix} 2 & -1 & -1 & -12 \\ -1 & 2 & -1 & -15 \end{pmatrix} \quad (5.6)$$

Having obtained \mathcal{T} , we attempt to get insight on how uncertainties in various input parameters propagate through the model causing uncertainty in the equilibrium quantities. If we assume that all these parameters, viz. γ_1, γ_2, a, b have a 10% coefficient of variation and are all uncorrelated, then the covariance matrix of the input is

$$C_1 = \begin{pmatrix} 0.04 & 0 & 0 & 0 \\ 0 & 0.01 & 0 & 0 \\ 0 & 0 & 2.25 & 0 \\ 0 & 0 & 0 & 0.01 \end{pmatrix} \quad (5.7)$$

Then the covariance matrix of the solution would be

$$C_1^* = \mathcal{T}C_1\mathcal{T}^T = \begin{pmatrix} 0.4289 & 0.4389 \\ 0.4389 & 0.5089 \end{pmatrix} \quad (5.8)$$

The standard deviation of the produced quantities are $0.65(= \sqrt{0.4289})$ and $0.71(= \sqrt{0.5089})$ respectively. The produced quantities also have about 95% positive correlation as an increase in demand will cause both producers to produce more and a decrease in demand will cause both producers to produce less.

If we assume that we have perfect knowledge about the demand curve, and if the uncertainty is only in the production costs, then the new parameter covariance C_2 has the third and fourth diagonal term of C_1 as zero. In such a scenario, we would expect the decrease in the quantity of production of one player to cause an increase in the quantity of production of the other and vice versa since one producer would lose market share to the other. We can see that by computing the covariance of the solution as $\mathcal{T}C_2\mathcal{T}$. The solution thus obtained shows that the produced quantities are negatively correlated with a correlation of -85% . The uncertainties in the produced quantities are 3% and 2% respectively of the quantity produced by each producer. We also note that the variances are smaller now, as we no longer have uncertainties stemming from the demand side of the problem.

Now if we assume a more realistic scenario of the production costs being correlated (60% correlation), then we note that the produced quantity are negatively correlated with -62% correlation. The standard deviations in the produced quantities have also dropped to about 2.9% and 1.2% of the produced quantities. Thus we not only obtain insight about the uncertainties in the output, but also the correlation between the output parameters. From an energy market policy maker's perspective this is crucial information as it helps identifying the regions where production/consumption/price/flows increase/decrease simultaneously and where they change asynchronously. Now we calculate the sensitivity of each of the input parameters to identify the parameter that causes maximum uncertainty in the output. The values for β for each of the four parameters γ_1, γ_2, a, b are calculated below.

$$\beta = \frac{1}{3} (\sqrt{5} \quad \sqrt{5} \quad \sqrt{2} \quad \sqrt{369})^T = (0.745 \quad 0.745 \quad 0.471 \quad 6.40)^T \quad (5.9)$$

Thus we see that the solution is more sensitive to the slope of the demand curve than to say production cost. Strictly speaking, this says, a unit increase in uncertainty (variance) of the slope of the demand curve will be magnified about 41 times (6.4^2) in the output uncertainty. However a unit increase in the uncertainty of production cost will increase the uncertainty in the equilibrium only by 0.556 units.

5.2 Computational Complexity

A Monte-Carlo simulation based approach has been chosen as a comparison to the approximation procedure discussed in this paper as this is the most commonly available method to calculate covariance in a complementarity problem. For the Monte-Carlo simulation based approach, a symmetrically balanced stratified design ([40]) is used with each dimension divided into two strata. With increasing number of random-parameters and equilibrium variables, Monte-Carlo methods become increasingly inefficient as the number of simulations required grows exponentially. A comparison of the time taken in an *8GB RAM 1600 MHz DDR3 2.5GHz Intel Core i5* processor to solve the above oligopoly problem with varying number of players is shown in Fig. 2. Despite developments in algorithms to solve complementarity problems, the said exponential growth in the number of sample points required in a Monte-Carlo based approach deters the computational speed. A problem with as few as 25 uncertain variables takes about 2 hours to solve and one with 30 uncertain variables takes about seven days to solve using Monte-Carlo based approaches.

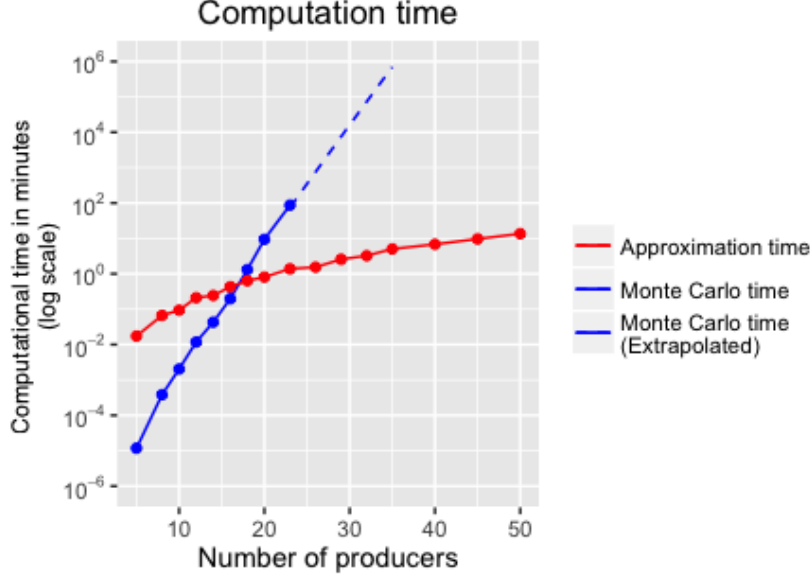


Figure 2: Time complexity of Monte-Carlo methods vs our First-order approximation method

6 Application to North American Natural Gas Market

In Mexico, motivation to move from coal to cleaner energy sources creates an increasing trend in natural gas consumption. United States is expected to become a net exporter of Natural gas by 2017 ([2]) especially due to the continuous increase in exports to Mexico ([17]). The North American Natural Gas Model (NANGAM) developed in [19] analyzes the impacts of cross border trade with Mexico. NANGAM models the equilibrium under various scenarios by competitively maximizing the profits of producers and pipeline operators and the utility of consumers. The formulation leads to a complementarity problem. The model also uses the Golombek function ([25]) in the supply function to model the increase in marginal cost of production, when producing close to capacity. This makes the complementarity problem into a nonlinear one. The formal description of the model is provided in Appendix A.

In this model, which is motivated by NANGAM, we have disaggregated United states into 9 census regions and Alaska ([2]). Mexico is divided into 5 regions. A map showing this regional disaggregation is shown in Fig. 5. Further Canada is divided into two zones, Canada East and Canada West. The model has 13 producers, 17 consumers, 17 nodes and 7 time-steps. This amounts to 12,047 variables (primal and dual) and 2023 parameters. The gradient matrix of the complementarity function would contain $12,047^2$ elements and a Hessian matrix will have 12047^3 elements which is more than 1700 trillion floating point variables. We need efficient methods to handle these large objects. We observe, however, that the dependence of each component of the complementarity function is limited to few variables, thus making the gradient matrix sparse. Efficient sparse matrix tools in scipy ([31]) are used along with a python class we specially built to handle a sparse multi-dimensional array. The details of this class are given in Appendix B.

This model is calibrated to match the region wise production and consumption data/projections by adjusting the parameters of the demand curve, supply curve and the transportation cost. The source for the projected numbers are the same as the ones in Table 2 of [19]. The parameters of the demand curve were chosen in such a way that an elasticity of 0.29 is maintained at the solution to be consistent with [16].

6.1 Covariance Matrix Calibration

We used the method developed in algorithm 1 to understand the propagation of uncertainty in the model. The covariance for each parameter across years is obtained by fitting a Wiener process to the parameter value. This is chosen to mimic the Markovian and independent increment properties of market

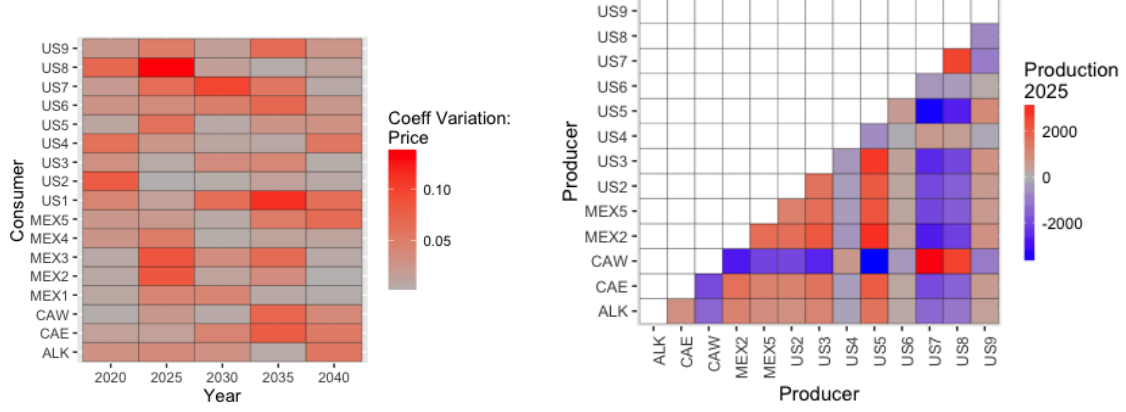


Figure 3: Coefficient of variation in Price and Covariance of Produced quantity

parameters. Thus we have for any parameter

$$d\theta(t) = d\mu_\theta(t) + \sigma_\theta dB(t) \quad (6.1)$$

where μ_θ is calibrated, σ_θ is chosen to be 1% of the average value of μ_θ in the analyzed period and $B(t)$ is the standard Brownian motion. The diffusion parameter σ_θ is assumed to be independent of time. Additionally to understand the effect of greater uncertainty in the US census region US7, which contains Texas and accounts for about 40% of the total production in the continent, the parameters of production cost are assumed to have 5 times the variance than in any other region.

6.2 Results

The deterministic version of the problem is solved using the PATH algorithm ([14]) by assuming a mean value for all random parameters. With that solution as the initial point, we used the algorithm in [30] to solve the stochastic version of the problem and obtain the solution to the SCP. Following this, algorithm 1 was applied and the \mathcal{T} matrix defined in (2.65) is obtained by solving the linear system of equations using a Moore Penrose pseudo inverse ([27]). In the following paragraph, we discuss some of the results obtained in this study.

The heat map on the left of Fig. 3 shows the coefficient of variation (std dev divided by mean) in consumer price in each year caused by the uncertainty in parameters as mentioned in subsection 6.1. We notice that this uncertainty in production costs of US7 caused relatively small uncertainties in the consumer price. This is partially due to the availability of resources in US8 and Canada West to compensate for the large uncertainty in Texas (US7). The fact that it is actually US8 and Canada West that compensate for this uncertainty is known by looking at the covariance plot on the right of Fig. 3 which shows large correlation between US7 and US8 and also between US7 and CAW.

Fig. 4 shows the sensitivity of the solution to various input parameters. The graph on the left shows the sum total change in uncertainty in price for a 1% fluctuation in the demand curve of various consumers. We notice that the price is particularly sensitive to changes in demand in Mexico. We also note that fluctuations in demand at nodes where production facilities are not available (MEX1, MEX3, MEX4) cause greater uncertainty in price. This is because, for regions with a production facility in the same node, the production facility produces more to cater the demand at that node and there is little effect in the flows and in the prices at other nodes. However a perturbation to the demand at a node with no production unit causes the flows to alter to have its demand catered. This affects natural gas availability elsewhere and causes larger fluctuations in price. The tornado plot on the right of Fig. 4 sorts the parameters in decreasing order of their effect on the uncertainty of the solution. Noting that it is plotted in log scale, we understand that uncertainties in the intercept of the demand curve affects the equilibrium the most. Among the supply parameters the linear cost parameter causes maximum fluctuation in the output.

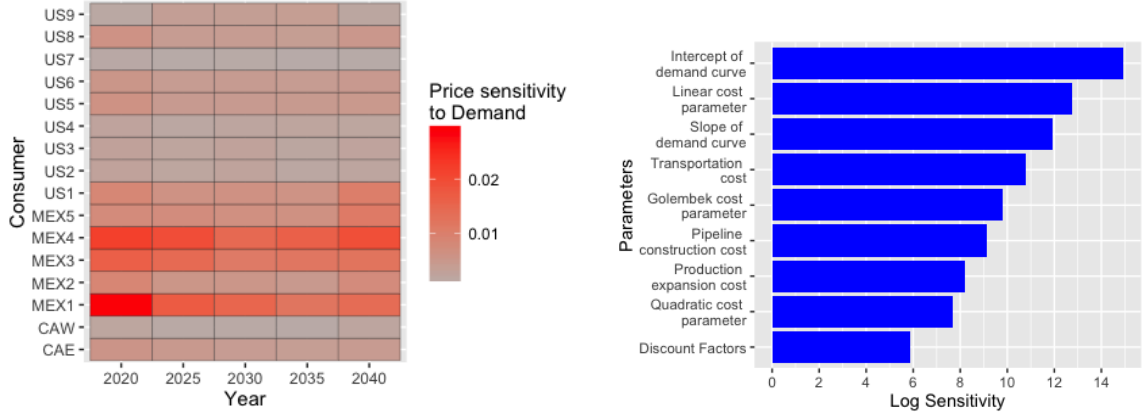


Figure 4: Sensitivity of the solution to the parameters

7 Conclusion and Future work

In this paper, we developed a method to approximate the covariance of the output of a large-scale nonlinear complementarity problem with random input parameters using first-order metrics. We extended this method to a general unconstrained optimization problem. We then developed sensitivity metrics for each of the input parameters quantifying their contribution to the uncertainty in the output. We used these tools to understand the covariance in the equilibrium of the North American Natural Gas Market. The method gave insights into how production, consumption, pipeline flows, prices would vary due to large uncertainties. While the variances identified the regions that are affected the most, the covariance gave information about whether the quantity will increase or decrease due to perturbation in the input. We also obtained results on the sensitivity of price uncertainty to demand uncertainty in various nodes. We then quantified the contribution of each input parameter to the uncertainty in the output. This in turn, helps in identifying the regions that can have large impacts on equilibrium.

We note that the method is particularly useful for large-scale nonlinear complementarity problems with a large number of uncertain parameters to make Monte-Carlo simulations or methods involving scenario trees intractable. It is robust in approximating the solution covariance for small uncertainty in the inputs. It is also good in quantifying the sensitivity of the output (and its variance) to the variance of various input parameters. However since all the above are obtained as an approximation based on first-order metrics, there is a compromise in the accuracy if the variances of the input are large. The method works the best for problems involving a large number of decision variables and random parameters with small variance.

We foresee expanding this work by using progressively higher-order terms of the Taylor series to capture the nonlinearities more efficiently. To ensure computational feasibility, this would typically require us to have stronger assumptions on the sparsity of the Hessian and the higher-order derivatives. This will also require knowledge/assumptions about higher-order moments of the random parameters.

8 Acknowledgements

The authors would like to thank Dr. Donniell Fishkind, Department of Applied Mathematics and Statistics, Johns Hopkins University, Dr. Michael Ferris, Department of Computer Sciences, University of Wisconsin-Madison and the participants of TAI conference 2016, MOPTA 2016 and INFORMS 2016 for their valuable comments and discussions.

A Natural Gas Market - Complementarity Formulation

In this formulation, we assume we have a set of producers P , consumers C and a pipeline operator. The players are located in a set of nodes N , and some of them are connected by pipelines A .

Let also say that $P_n \subseteq P$, $C_n \subseteq C$ are located in node $n \in N$. Let A_n be the pipelines connected to node

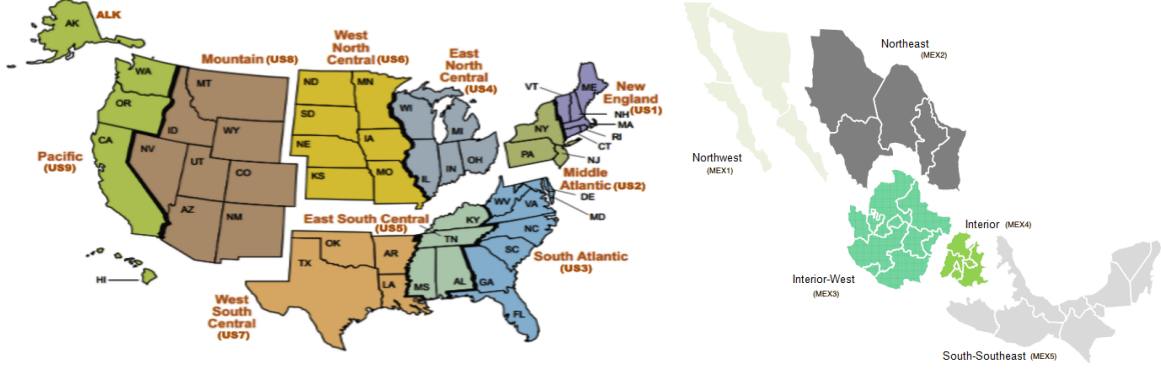


Figure 5: Regional disaggregation of United States and Mexico. Source: [2] and U.S. Energy Information Administration <http://www.eia.gov/todayinenergy/detail.php?id=16471>

n . The symbols used here are explained in Table 1, 2 and 3. Most of the analysis closely follow [19] and [15]. Random parameters are denoted by an (ω) beside them. The implementation of this problem is made available in https://github.com/ssriram1992/Stoch_Aprx_cov.

A.1 Producer's problem

$$\begin{aligned} \text{Maximize } \sum_Y df_y(\omega) \left\{ \sum_C Q_{pcny}^C \pi_{cy} - \text{Gol}(Q_{pny}^P, CAP_{py}^P) \right. \\ \left. - \pi_{py}^{XP}(\omega) X_{py}^P - \sum_{A_n^+} \pi_{ay}^A Q_{pay}^A \right\} \end{aligned} \quad (\text{A.1})$$

subject to

$$\begin{aligned} Q_{pcny}^C, Q_{pny}^P, Q_{pay}^A &\geq 0 \\ X_{py}^P, CAP_{py}^P &\geq 0 \\ Q_{pny}^P &\leq \alpha^P CAP_{py}^P \end{aligned} \quad (\delta_{py}^1) \quad (\text{A.2a})$$

$$CAP_{py}^P = \hat{Q}_{p0} + \sum_{i=1}^y X_{pi}^P \quad (\delta_{py}^2) \quad (\text{A.2b})$$

$$\begin{aligned} \sum_{C_n} Q_{pcny}^C + \sum_{A_n^+} Q_{pay}^A &= Q_{pny}^P (1 - L_{py}^P(\omega)) \\ &+ \sum_{A_n^-} Q_{pay}^A (1 - L_{ay}^A(\omega)) \end{aligned} \quad (\delta_{pny}^3) \quad (\text{A.2c})$$

where

$$\begin{aligned} \text{Gol}(\cdot) &= (l_{py}^P(\omega) + g_{py}^P(\omega)) Q_{pny}^P + q_{py}^P(\omega) Q_{pny}^P{}^2 \\ &+ g_{py}^P(\omega) (CAP_{py}^P - Q_{pny}^P) \log \left(1 - \frac{Q_{pny}^P}{CAP_{py}^P} \right) \end{aligned} \quad (\text{A.3})$$

A.2 Pipeline operator's problem

$$\text{Maximize } \sum_Y df_y(\omega) \left\{ \sum_A Q_{ay}^A (\pi_{ay}^A - \gamma_{ya}^A(\omega)) - \pi_{ay}^{XA}(\omega) X_{ay}^A \right\} \quad (\text{A.4})$$

Table 1: Sets

Set	Explanation	Set	Explanation
P	Set of producers	A	Set of pipeline connections(arcs)
C	Set of consumers	A_n^+	Set of arcs from node n on which natural gas flows out
N	Set of nodes	A_n^-	Set of arcs from node n on which natural gas flows in
Y	Set of periods		

Table 2: Symbols - Variables

	Symbol	Explanation
Quantities	Q_{pcny}^C	Quantity produced by p in n to send to c in year y
	Q_{pny}^P	Total quantity produced by p in year y
	Q_{pay}^A	Total quantity p choses to send by arc a in year y
	Q_{ay}^A	Total quantity sent by a during year y
Prices	π_{cy}	Unit price paid by consumer C in year Y
	π_{ay}^A	Unit price of sending natural gas through a during year y
Capacity	X_{py}^P	Production expansion in year y for producer p
	X_{ay}^A	Transportation capacity expansion in year y for arc a
	CAP_{py}^P	Production capacity for producer p in year y
	CAP_{ay}^A	Transportation capacity for arc a in year y

Table 3: Symbols - Parameters

	Symbol	Explanation
Quantities	\hat{Q}_{p0}	Initial capacity of production for producer p
	\hat{Q}_{a0}	Initial capacity of transportation for pipeline a
Prices	$\pi_{py}^{XP}(\omega)$	Price of capacity expansion for producer p
	$\pi_{ay}^{XA}(\omega)$	Price of capacity expansion for transportation arc a
Losses	$L_{py}^P(\omega)$	Percentage loss in production by producer p in year y
	$L_{ay}^A(\omega)$	Percentage loss in transportation via arc a in year y
	α^P	Availability fraction of the production capacity
Consumer	$E_{cy}^C(\omega)$	Intercept of the demand curve for consumer c in year y
	$D_{cy}^C(\omega)$	Slope of the demand curve for consumer c in year y
	$df_y(\omega)$	Discount Factor for year y

subject to

$$\begin{aligned} Q_{ay}^A, X_{ay}^A, CAP_{ay}^A &\geq 0 \\ Q_{ay}^A &\leq CAP_{ay}^A \end{aligned} \quad (\delta_{ay}^5) \quad (A.5a)$$

$$CAP_{ay}^A = \hat{Q}_{a0} + \sum_{i=1}^y X_{ai}^A \quad (\delta_{ay}^6) \quad (A.5b)$$

A.3 Consumer

$$\pi_{cy} = E_{cy}^C(\omega) + D_{cy}^C(\omega) \sum_P Q_{pcny}^C \quad (\pi_{cy}) \quad (A.6)$$

It can be shown that the above said optimization problems are all convex with non-empty interior. Hence the Karush-Kuhn Tucker conditions (KKT conditions) are necessary and sufficient for optimality. The KKT conditions are presented below and they form the equations for the complementarity problem along with the constraints above.

A.4 KKT to Producer's problem

$$-df_y(\omega)\pi_{cy} + \delta_{pny}^3 \geq 0 \quad (Q_{pcny}^C) \quad (A.7a)$$

$$df_y(\omega)\pi_{py}^{XP}(\omega) - \sum_{i=1}^y \delta_{pi}^2 \geq 0 \quad (X_{py}^P) \quad (A.7b)$$

$$df_y(\omega)\pi_{ay}^A + \left(\mathbb{I}_{a \in A_n^+} - \mathbb{I}_{a \in A_n^-} (1 - L_{ay}^A(\omega)) \right) \delta_{pny}^3 \geq 0 \quad (Q_{pay}^A) \quad (A.7c)$$

$$df_y(\omega) \frac{\partial \text{Gol}}{\partial Q_{pny}^P} + \delta_{py}^1 - \delta_{pny}^3 (1 - L_{py}^P(\omega)) \geq 0 \quad (Q_{pny}^P) \quad (A.7d)$$

$$df_y(\omega) \frac{\partial \text{Gol}}{\partial CAP_{py}^P} + \alpha^P \delta_{py}^2 - \delta_{py}^1 \geq 0 \quad (CAP_{py}^P) \quad (A.7e)$$

A.5 KKT to Pipeline operator's problem

$$-df_y(\omega)\pi_{ay}^A + \gamma_{ya}^A(\omega) + \delta_{ay}^5 \geq 0 \quad (Q_{ay}^A) \quad (A.8a)$$

$$df_y(\omega)\pi_{ay}^{XA}(\omega) - \sum_{i=1}^y \delta_{ai}^6 \geq 0 \quad (X_{ay}^A) \quad (A.8b)$$

$$\delta_{ay}^6 - \delta_{ay}^5 \geq 0 \quad (CAP_{ay}^A) \quad (A.8c)$$

A.6 Market clearing condition

$$Q_{ay}^A = \sum_P Q_{pay}^A \quad (\pi_{ay}^A) \quad (A.9)$$

B N-dimensional Sparse array implementation

A general purpose Python class has been implemented to handle a sparse *ndarray* object. The class is a generalization of the *scipy* class *coo_matrix* which stores the array coordinates of each non-zero element in the array. We now describe the details of the implementation. A continuously updated version of the class can be found at <https://github.com/ssriram1992/ndsparse>.

B.1 Initialization

The n-dimensional sparse array (`coo_array`) can be initialized by any of the following methods.

- A *tuple*, which initializes the sparse array of the shape mentioned in the *tuple* and with zeros everywhere.
- A dense *ndarray* which will be converted and stored as a `coo_array`.
- A *matrix* of positions and a 1 dimensional *array* of values where the matrix contains the positions of the non-zero elements and the vector containing the non-zero values of those positions. In this case the shape of the `coo_array` would be the smallest *ndarray* that can store all the elements given. Optionally a *tuple* containing the shape of the *ndarray* can be given explicitly.
- Another *coo_array* whose copy is to be created.

B.2 Methods

The following methods and attributes are available in the `coo_array`.

- `print(coo_array)` will result in printing the location of each of the non-zero elements of the array and their values.
- `coo_array.flush(tol = 1e-5)` will result in freeing the space used in storing any zero-elements or elements lesser than the tolerance, `tol`. Such numbers typically arise out arithmetic operations on `coo_array` or poor initialization.
- `coo_array.size()` returns the number of non-zero elements in the `coo_array`.
- `coo_array.shape` returns the shape of the underlying dense matrix.
- `coo_array.add_entry(posn, val)` and `coo_array.set_entry(posn, val)` both add a new non-zero element with the given value at the given position. The difference however is that `set_entry()` checks if a non-zero value already exists at the mentioned position, and if yes, overwrites it. This search makes `set_entry()` slower compared to `add_entry()` which assumes that the previous value or the position is zero. Thus `add_entry()` could potentially cause duplicates and ambiguity, if an illegal input is given. However in case the input is ensured to be legal, `add_entry()` is much faster.
- `coo_array.get_entry(posn)` returns the value at the given position.
- `coo_array.swapaxes(axis1, axis2)` is a higher dimensional generalization of matrix transposes where the dimensions that have to swapped can be chosen.
- `coo_array.remove_duplicate_at(posn, func=0)` checks if there are multiple values defined for a single position in the sparse array. If yes, they are replaced by a single entry containing the scalar valued defined by `func` or passes them to a function defined in `func` and stores the returned value. Passing a function for the argument `func` is incredibly useful in performing arithmetic operations on `coo_array`.
- `coo_array.todense()` returns a dense version of the `coo_array`.
- `coo_array.iterate()` returns an iterable over the non-zero positions and values in the `coo_array`.

The above class is used extensively to handle high-dimensional sparse arrays resulting out of variables containing pipelines, viz., $Q_{pay}^A, Q_{ay}^A, \hat{Q}_{p0}, \hat{Q}_{a0}, \pi_{ay}^A, X_{ay}^A, \delta_{ay}^5, \delta_{ay}^6$ and parameters with pipelines, viz., $CAP_{ay}^A, \pi_{ay}^{XA}(\omega), L_{ay}^A(\omega), \gamma_{ya}^A(\omega)$.

References

- [1] Ibrahim Abada, Steven Gabriel, Vincent Briat, and Olivier Massol. A generalized nash–cournot model for the northwestern european natural gas markets with a fuel substitution demand function: The gammes model. *Networks and Spatial Economics*, 13(1):1–42, 2013.
- [2] Annual Energy Outlook AEO. US Energy Information Administration, 2015, 2015.
- [3] Shipra Agrawal, Y. Ding, A. Saberi, and Y. Ye. Price of Correlations in Stochastic Optimization. *Operations Research*, 60(1):150–162, 2012.
- [4] Edward J Anderson and Huifu Xu. Nash equilibria in electricity markets with discrete prices. *Mathematical Methods of Operations Research*, 60(2):215–238, 2004.
- [5] Aloisio Araujo and Evarist Giné. *The central limit theorem for real and Banach valued random variables*, volume 431. Wiley New York, 1980.
- [6] Mor Armony and Avishai Mandelbaum. Routing and Staffing in Large-Scale Service Systems: The Case of Homogeneous Impatient Customers and Heterogeneous Servers. *Operations Research*, 59(1):50–65, 2011.
- [7] Lisandro Benedetti-Cecchi. The importance of the variance around the mean effect size of ecological processes. *Ecology*, 84(9):2335–2346, 2003.
- [8] Bahar Biller and Canan G Corlu. Accounting for Parameter Uncertainty in Large-Scale Stochastic Simulations with Correlated Inputs. *Operations Research*, 59(3):661–673, 2011.
- [9] James Bushnell. A mixed complementarity model of hydrothermal electricity competition in the western United States. *Operations research*, 51(1):80–93, 2003.
- [10] Xiaojun Chen and Masao Fukushima. Expected Residual Minimization Method for Stochastic Linear Complementarity Problems. *Mathematics of Operations Research*, 30(4):1022–1038, 2005.
- [11] Adam Christensen and Sauleh Siddiqui. A mixed complementarity model for the us biofuel market with federal policy interventions. *Biofuels, Bioproducts and Biorefining*, 9(4):397–411, 2015.
- [12] Ivan Contreras, Jean-François Cordeau, and Gilbert Laporte. Benders Decomposition for Large-Scale Uncapacitated Hub Location. *Operations Research*, 59(6):1477–1490, 2011.
- [13] Richard W Cottle, Jong Shi Pang, and Richard E Stone. *The linear complementarity problem*, volume 60. Siam, 2009.
- [14] Steven P Dirkse and Michael C. Ferris. The path solver: a nommonotone stabilization scheme for mixed complementarity problems. *Optimization Methods and Software*, 5(2):123–156, 1995.
- [15] Ruud Egging, Alois Pichler, Øyvind Iversen Kalvø, and Thomas Meyer WalleHansen. Risk Aversion in Imperfect Natural Gas Markets. *European Journal of Operational Research*, 2016.
- [16] Energy Information Administration EIA. Fuel Competition in Power Generation and Elasticities of Substitution, 2012.
- [17] Energy Information Administration EIA. U.s. natural gas exports to mexico continue to grow, 2016.
- [18] Francisco Facchinei and Jong Shi Pang. *Finite-dimensional variational inequalities and complementarity problems*. Springer Science & Business Media, 2007.
- [19] Felipe Feijoo, Daniel Huppmann, Larissa Sakiyama, and Sauleh Siddiqui. North American natural gas model: Impact of cross-border trade with Mexico. *Energy*, 112:1084–1095, oct 2016.
- [20] Michael C. Ferris and Jong Shi Pang. Engineering and Economic Applications of Complementarity Problems. *SIAM Review*, 39(4):669–713, 1997.

- [21] Steven A Gabriel, Antonio J Conejo, J David Fuller, Benjamin F Hobbs, and Carlos Ruiz. *Complementarity modeling in energy markets*, volume 180. Springer Science & Business Media, 2012.
- [22] Steven A Gabriel, Andy S Kydes, and Peter Whitman. The National Energy Modeling System: A Large-Scale Energy-Economic Equilibrium Model. *Operations Research*, 49(1):14–25, 2001.
- [23] Steven A. Gabriel, Jifang Zhuang, and Ruud Egging. Solving stochastic complementarity problems in energy market modeling using scenario reduction. *European Journal of Operational Research*, 197(3):1028–1040, 2009.
- [24] J. Galambos. *The asymptotic theory of extreme order statistics*. Wiley series in probability and mathematical statistics. Applied probability and statistics. Wiley, 1978.
- [25] Rolf Golombek, Eystein Gjelsvik, and Knut Einar Rosendahl. Effects of liberalizing the natural gas markets in Western Europe. *The Energy Journal*, pages 85–111, 1995.
- [26] Gül Gürkan and Stephen M. Robinson. Sample-path solution of stochastic variational inequalities. *Mathematical Programming, Series B*, 84(2):313–333, 1999.
- [27] Roger A Horn and Charles R Johnson. *Matrix analysis*. Cambridge university press, 2012.
- [28] Daniel Huppmann and Ruud Egging. Market power, fuel substitution and infrastructure—a large-scale equilibrium model of global energy markets. *Energy*, 75:483–500, 2014.
- [29] Kate L Hyett, Mike Podosky, Nick Santamaria, and Jenni C Ham. Valuing variance: the importance of variance analysis in clinical pathways utilisation. *Australian Health Review*, 31(4):565–570, 2007.
- [30] Houyuan Jiang and Huifu Xu. Stochastic approximation approaches to the stochastic variational inequality problem. *IEEE Transactions on Automatic Control*, 53(6):1462–1475, 2008.
- [31] Eric Jones, Travis Oliphant, Pearu Peterson, and Others. {SciPy}: Open source scientific tools for {Python}, 2015.
- [32] Byung Man Kwak and Seok Soon Lee. A complementarity problem formulation for two-dimensional frictional contact problems. *Computers & structures*, 28(4):469–480, 1988.
- [33] Michael Lamm, Shu Lu, and Amarjit Budhiraja. Individual confidence intervals for solutions to expected value formulations of stochastic variational inequalities. *Mathematical Programming*, pages 1–46, jul 2016.
- [34] Jun Luo, L Jeff Hong, Barry L Nelson, and Yang Wu. Fully Sequential Procedures for Large-Scale Ranking-and-Selection Problems in Parallel Computing Environments. *Operations Research*, 63(5):1177–1194, 2015.
- [35] Sebastián Martín, Yves Smeers, and José Antonio Aguado. A Stochastic Two Settlement Equilibrium Model for Electricity Markets With Wind Generation. *IEEE Transactions on Power Systems*, 30(1):1–13, 2015.
- [36] Jorge Nocedal and Stephen Wright. *Numerical optimization*. Springer Science & Business Media, 2006.
- [37] Olufolajimi Oke, Daniel Huppmann, Max Marshall, Ricky Poulton, and Sauleh Siddiqui. Mitigating environmental and public-safety risks of united states crude-by-rail transport. 2016.
- [38] George A F Seber and Alan J Lee. *Linear regression analysis*, volume 936. John Wiley & Sons, 2012.
- [39] Uday V. Shanbhag. *Stochastic Variational Inequality Problems : Applications , Analysis , and Algorithms*. INFORMS, 2013.
- [40] Michael D Shields, Kirubel Teferra, Adam Hapij, and Raymond P Daddazio. Refined stratified sampling for efficient monte carlo based uncertainty quantification. *Reliability Engineering & System Safety*, 142:310–325, 2015.

- [41] Sauleh Siddiqui and Adam Christensen. Determining energy and climate market policy using multiobjective programs with equilibrium constraints. *Energy*, 94:316–325, 2016.
- [42] Justin A. Sirignano, Gerry Tsoukalas, and Kay Giesecke. Large-scale loan portfolio selection. *Operations Research*, 64(6):1239–1255, 2016.
- [43] Amy R Ward and Mor Armony. Blind Fair Routing in Large-Scale Service Systems with Heterogeneous Customers and Servers. *Operations Research*, 61(1):228–243, 2013.
- [44] Paul Wilmott, Jeff Dewynne, and Sam Howison. *Option pricing*. Oxford Financial Press, 1994.